MANISH SHRIVASTAV

manishshrivastav@live.com
https://www.linkedin.com/in/manishshrivastav83/
515-720-5811
https://github.com/ManishShrivastav

DATA SCIENCE | MACHINE LEARNING

MOTIVATION

*I am passionate about **solving business problems** using Data Science & Machine Learning. I systematically & creatively use my skillset to **add tangible value** to the team, the business, and the end-user. I am constantly learning, and always looking to improve.*

SKILLS & TOOLS

Programming: SQL, Python (Base, Tensorflow, PyTorch, Pandas, Numpy, Matplotlib, Scikit-Learn, Keras)
Tools: Excel, Tableau, Github, AWS (S3, Lambda, IAM, EC2, SageMaker, RDS)
Math: Linear Algebra, Statistics (Hypothesis Testing, AB Testing, Central Limit Theorem, Distributions)
Machine Learning: Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis, Neural Networks

EXPERIENCE

Software Engineer - Corteva
May 2023 - sep 2024

- **Developed and optimized web platforms** that captured and processed over **50,000 acres** of carbon farming data annually, improving the integration of field data for accurate carbon credit calculations.
- **Designed and implemented RESTful APIs**, facilitating secure data exchange between data sources including soil sampling results and carbon credit metrics, enabling the system to handle **millions of data points** efficiently.
- **Implemented machine learning models** that predicted carbon sequestration patterns, improving carbon credit forecasting accuracy by **15%** based on historical data, soil conditions, and environmental variables.
- **Collaborated with cross-functional teams** to design data pipelines that tracked and **reported over 100,000 farming practices** across multiple regions, ensuring data integrity for carbon credit generation certified by Climate Action Reserve.

Data QA Engineer - Corteva
DEC 2022 - MAY 2023

- **Conducted exploratory data analysis (EDA) on 10+ years of carbon data**, identifying key trends and improving carbon sequestration model predictions by 18%.
- Applied **statistical methods** to analyze over 1,000 records, resolving data inconsistencies that could have decreased carbon credit generation by **10%**.
- Collaborated with data scientists to **refine data collection methods**, resulting in a **20% improvement** in carbon credit yield forecasting.

Machine Learning Engineer – Benekiva
JAN 2021 - FEB 2022

- **Extracted and analyzed data from death certificates** using Google Cloud Platform (GCP) to identify key information such as names, dates of birth, dates of death, causes of death, and familial relationships, **improving data retrieval efficiency by 30%**.
- **Developed real-time reporting tools** to help the client visualize trends in cause-of-death data, improving **decision-making** for the design of new life insurance products based on regional and demographic health risks.
- Enhanced **fraud detection** by cross-referencing death certificate data with internal policyholder records, ensuring claims were legitimate and reducing false claims by 5**%**.
- **Automated report generation and data analysis** processes, cutting the time required for death certificate review by **15 hours per week**, and improving operational efficiency.

PROJECTS

"You Are What You Eat" Customer Segmentation

- Used **k-means clustering** on grocery transaction data to split out customers into distinct "shopper types" that could be used to better understand customers over time, and to more accurately target customers with relevant content & promotions

Assessing Campaign Performance Using Chi-Square Test for Independence

- Performed **A/B testing** with a **Chi-Square Test** for Independence to assess the impact of two types of promotional mailers on customer signup rates for a grocery retailer's new "Delivery Club" service. Analyzed campaign data using Python to determine if there was a significant difference between the low-cost and high-cost mailers' effectiveness.

Quantifying Sales Uplift With Causal Impact Analysis

- Used **Causal Impact Analysis** to evaluate the sales uplift from customers who joined a grocery retailer's "Delivery Club" campaign. By comparing sales data from club members to a control group of non-members, the analysis revealed a significant increase in spending, indicating a significant causal effect of the campaign. This insight helps measure the campaign's effectiveness and can inform future marketing strategies.

Understanding Product Relationships Using Association Rule Learning

- Utilized **Association Rule Learning with the Apriori algorithm** to analyze transactional data and uncover product relationships. The goal was to provide actionable insights for optimizing product placement, recommending bundled promotions, and improving the customer shopping experience.

Compressing Feature Space for Classification Using PCA

- Applied **Principal Component Analysis (PCA)** to reduce a high-dimensional, sparse feature set into a more manageable form, enabling efficient classification of customers who purchased Ed Sheeran's previous album. The goal was to build a predictive model using compressed features to help the client target marketing efforts for future album promotions.

EDUCATION

Ph.D. (Agricultural Engineering, Statistics-Minor)
2016-2022 – Iowa State University, IA

COURSES & CERTS

Data Science Professional Certification (Data Science Infinity)
Actionable Learnings: Extracting & manipulating data using SQL. Application of statistical concepts such as hypothesis tests for measuring the effect of AB Tests. Utilising Github for version control, and collaboration. Using Python for data analysis, manipulation & visualisation. Applying data preparation steps for ML including missing values, categorical variable encoding, outliers, feature scaling, feature selection & model validation. Applying Machine Learning algorithms for regression, classification, clustering, association rule learning, and causal impact analysis for measuring the impact of an event over time. Machine Learning pipelines to streamline the ML pre-processing & modelling phase. Deployment of a ML pipeline onto a live website using Streamlit. Using Tableau to create powerful Data Visualizations. Turning business problems into Data Science solutions.